

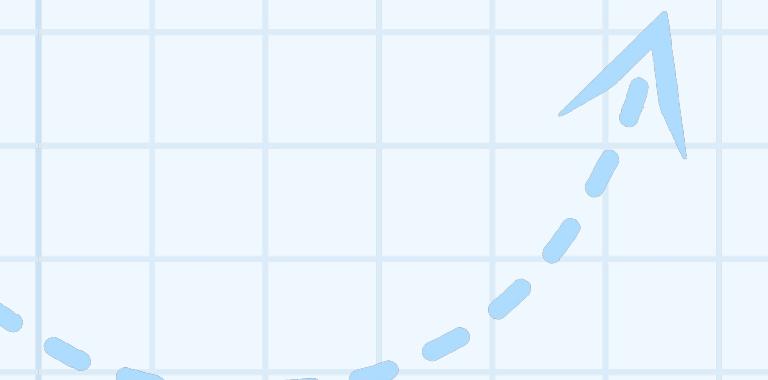
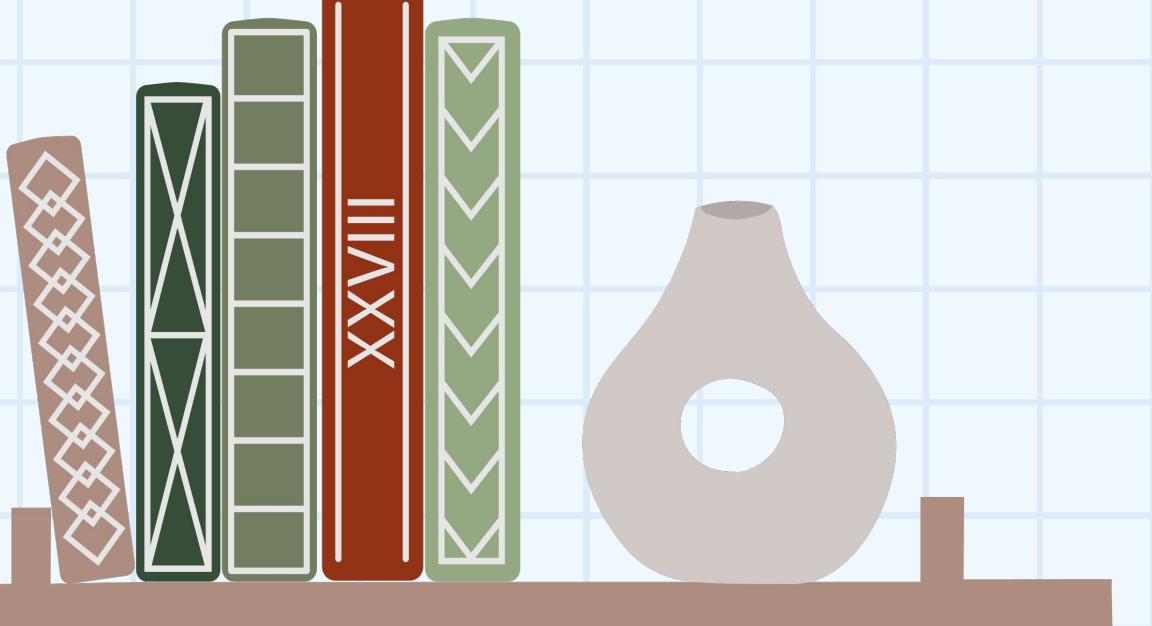
Powered by
Futurense



BS./BSC.IN

Applied AI and Data Science

Basics of Data Analytics





Powered by

Futurense

Let's dive into and learn:



1

Identifying Data



Identifying the data

1. Determine the information you want to collect

- Linked closely to the main question or problem statement
 - Specific information you need
 - Sources of the information
 - Outline a plan for collecting the data



Identifying the data

2. Outline a plan for collecting the data

- Decide on the timeline for collecting data
- E.g. for data from web platforms, may need to refresh at intervals
- How much data is needed? – sample size, power analysis



Identifying the data

3. Decide on the data collection methods. Depends on

- Sources of data
- Types of data
- Format of data
- Timeline
- Volume of data

Example



Powered by

Futurense

Suppose a retail company wants to create a targeted ad campaign to increase their sales ahead of the festive season.

You are the data analyst for this company and you need to help them do this.

What data do you think you would need?



Example

1. Determine the information you want

- Customer Profile – purchase history, age, gender, location, education or marital status
- Website visits and views
- Customer satisfaction survey responses
- Customer complaints
- Social media conversations about your products



Example

2. Timeline of collection

- Customer profile data – internal to the company
- Website visits and views – continuously refreshed
- Customer satisfaction surveys – design your own survey and sampling strategy
- Customer complaints – updated on an ongoing basis
- Social media feed – continuously updated

Example



Powered by

Futurense

3. Collection Methods

- Customer profile data – structured relational database, querying
- Website visits and views – data streams
- Customer satisfaction surveys – design your own survey and sampling strategy
- Customer complaints – text data, NLP
- Social media feed – APIs

Recap



Powered by
Futurense

How to Identify the relevant data for a given data
analytics project



Powered by



Futurense

Thank you

